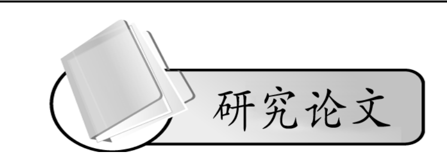


面向光伏项目投资风险的大数据监测指标甄选研究*

——以 Solarbao 平台为例

杨 旸[1] 林 辉[1] 胡广伟[2]

[1](南京大学商学院 南京 210093)

[2](南京大学信息管理学院 南京 210093)

摘要:【**目的**】在构建光伏项目投资风险监测模型的过程中，为了甄选面向互联网金融平台的大数据应用监测指标，尝试提出系统的甄选方案并结合实际案例进行验证。【**方法**】应用大数据监测模型，整合 Solarbao 平台多源异构数据，以专家判断为项目投资风险分析依据，运用 CHAID 决策树归纳多维监测指标组合，并运用 R-Q 型因子分析方法提炼识别投资风险的关键指标。【**结果**】得到 8 条监测光伏项目投资风险的指标组合和 10 项识别投资风险的关键指标。【**局限**】R-Q 型因子分析中的专业指标有待进一步细分并形成动态更新机制。【**结论**】该甄选方案能够满足大数据监测模型对指标采集的要求，对投资者评估光伏项目风险、平台筛选合适项目以及监管部门排查该领域系统性风险具有借鉴意义。

关键词: 大数据监测指标　光伏项目　投资风险　CHAID 决策树　R-Q 型因子分析

分类号: F276.3　F830　TM62

1 引 言

随着中国能源互联网国家战略的实施，以及电力工业绿色转型方案的逐步出台，电力系统在新能源、新材料、新储能技术等方面将加大科技创新和资金投入力度，促进电力工业绿色转型和低碳发展[1]。光伏发电项目作为新能源发电产业的新生力量，具有无噪声、无污染、安全可靠等优点。但因其发电成本高，光电转换效率和能量密度低，且受气候环境影响大等特点[2]，导致项目投资风险偏高，在大规模推广和向传统金融机构融资的过程中遇到了诸多阻力[3]。为此，电力新能源项目的融资渠道正逐渐转向社会资本，并促成电力新能源领域互联网金融平台的相继产生和快速发展[4]。

该领域互联网金融平台的快速发展促进了光伏项目资金融通的效率，但其背后积累的融资主体信用风险也日趋显现[4]。平台中资信状况较差的光伏项目一旦发生债务违约，将造成投资者严重的经济损失[5-6]；同时，由于平台融合了金融行业和互联网行业的高风险属性，项目发生违约后，其风险扩散速度更快捷、交叉传染更严重，平台的整体投资风险将随之上升[7]，进而会影响该领域互联网金融体系的安全稳定。因此，

通讯作者：胡广伟，ORCID：0000-0003-1303-363X，E-mail：hugw@nju.edu.cn。

*本文系国家电网科技开发项目“提升电力营销服务能力的大数据关键技术研究”(项目编号：SGTYHT/14-JS-188)、国家自然科学基金面上项目“双维度流动性调整的期权定价模型研究”(项目编号：71271110)、江苏省“六大人才高峰”项目“政务大数据资源开发技术与实现方法研究”(项目编号：2015-XXRJ-001)和中国经济改革研究基金会课题“互联网金融的风险与监管制度研究”的研究成果之一。

加强光伏项目投资风险监测成为推进该领域互联网金融持续健康发展的客观需要和当务之急。

面对海量、与日俱增和多源异构的风险关联数据，传统风险监测模型缺乏实时捕捉、管理和处理相应数据的能力，需要借助大数据技术进行改进，实现多源异构数据提取、存储和多条件快速组合查询技术[8]，形成光伏项目投资风险大数据监测模型。由于光伏项目投资风险的构成较为复杂，不仅要借助关键指标，还要通过综合多个指标和特征参量形成风险判断[9-10]；并且平台数据的非结构化特点要求监测模型不能采用固定不变的指标，而是需要根据平台实际发布的融资信息科学制定出反映项目投资风险的监测指标。因此，合适的监测指标甄选方案有待被提出，这也是构建大数据监测模型的核心难点。

本研究基于 Solarbao 平台①，尝试提出面向光伏项目投资风险的大数据监测指标甄选方案。在整合 Solarbao 平台多源异构数据的基础上，以专家判断为项目投资风险分析的依据，运用 CHAID 决策树归纳多维监测指标组合，并运用 R-Q 型因子分析方法提炼识别投资风险的关键指标，以满足光伏项目投资风险大数据监测模型对指标采集的要求。研究的创新之处在于：

(1) 提出甄选线上光伏项目投资风险监测指标的系统方案；

(2) 编制应用于提炼多维监测指标组合和关键指标的调查问卷样例；

(3) 基于平台多源异构数据，构建符合大数据监测模型指标采集要求的实验指标；

(4) 采取的数据挖掘应用技术具备提炼风险判断、处理多维数据的能力，并且模型输出结果易于解释。

2 光伏项目投资风险大数据监测模型的构建

大数据监测模型的应用环境是甄选监测指标的重要基础[11]，本研究以 Solarbao 平台作为监测模型的应用环境。Solarbao 平台是以实物融资租赁为商业模式的互联网创新金融服务平台，因其专门从事光伏项目融资租赁且平台中投资项目众多，故选择该平台进行研究。该平台兼具众筹和融资租赁的特点，投资者通过平台购买太阳能电池板，并将电池板委托融资租赁公司出租给发电企业使用，发电企业用发电所得电费和政府补贴支付使用电池板的租金。在整个过程中，客户无须接触到太阳能电池板，所得的回报是分期拿到的租金，即先期投入资金，后期获取收益，平台因此具有了互联网金融的属性。光伏项目投资风险大数据监测模型应用环境如图 1 所示：

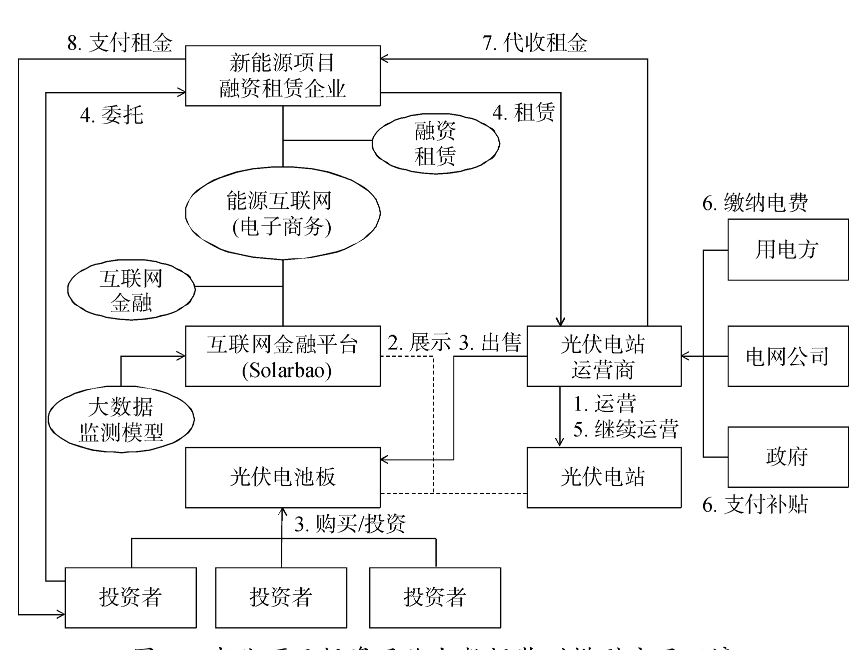


图 1 光伏项目投资风险大数据监测模型应用环境

①http://www.solarbao.com/.

由图 1 可知, Solarbao 平台是光伏项目和投资者之间的信息媒介, 融资标的信息、光伏电站运营商信息、光伏电站经营状况、政府补贴情况等将通过平台对外展示, 揭示该光伏项目开展融资和后期运营的基本条件, 为光伏项目投资风险监测指标甄选提供依据, 也形成了大数据监测模型的数据基础。光伏项目投资风险大数据监测模型如图 2 所示:

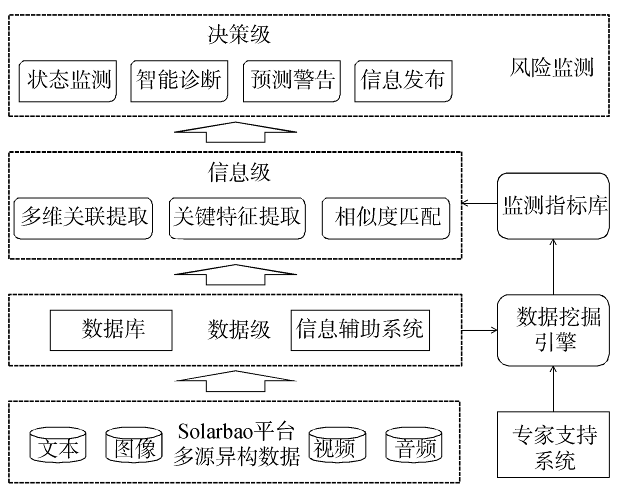


图 2　光伏项目投资风险大数据监测模型

由图 2 可知, 该模型具有分层信息聚合架构, 基于此架构进行光伏项目投资风险监测包含 4 个步骤:

(1) 数据提取和预处理。运用基于 GoldenGate 的多源异构数据提取方法[10], 从平台在线日志中实时抓取数据, 并将其保存在 Trail 格式的文件中; 数据遗漏、出现异常值等问题也在该步骤中进行处理。

(2) 数据级聚合。利用 Hadoop 的 MapReduce 并行计算框架加速装载抽取出的数据[10], 该框架自动对数据进行汇聚和排序, 将最终结果输出到 SQL Server 数据库和信息辅助系统中, MapReduce 并行计算框架提升了数据装载速度, 为大数据监测模型实时、动态获取数据提供技术保障。

(3) 信息级聚合。运用数据挖掘技术实时、动态甄选光伏项目投资风险监测指标, 呈现风险监测指标的多维关联和关键特征, 并提出与特定风险类型进行相似度匹配的方案。以往研究中关于项目风险判断的数据挖掘技术主要包括关联规则[12-13]、Kano 模型[14-15]、Kansei 工程[16]、CHAID 决策树[17-18]、R-Q 型因子分析等[19-20], 各数据挖掘技术的特点比较如表 1 所示。

表 1　数据挖掘技术特点比较

对比项目	关联规则	Kano 模型	Kansei 工程	CHAID 决策树	R-Q 型因子分析
提炼风险判断	不能	有限	能	能	能
提取风险偏好	能	能	不能	不能	能
处理多维数据	能	能	不能	能	能
解释的复杂度	高	低	高	低	低

由表 1 可知, 关联规则不适用于提炼风险判断且模型解释的复杂度较高; Kano 模型更侧重于提取客户对项目风险的偏好; Kansei 工程缺乏处理多维数据的能力; CHAID 决策树和 R-Q 型因子分析则相对适用, 可将专家判断作为项目投资风险分析的依据, 运用 CHAID 决策树实时、动态归纳多维监测指标组合, 并运用 R-Q 型因子分析方法实时、动态提炼识别投资风险的关键指标。

HBase 具备分布式并发处理效率高、易扩展、动态伸缩的特征[10], 因此本研究将数据挖掘的基础数据以一维或多维形式存储于 Hbase 中, 一方面支持数据挖掘引擎实时、动态分析数据, 另一方面响应数据监测中多条件快速组合查询的需求; 并通过 Hive 对基础数据进行备份, 便于后期离线计算和指标优化。

(4) 决策级聚合。为 HBase 中监测指标建立索引机制, 在数据导入时即通过索引引擎实现对监测数据低延迟、动态地提取, 进而通过 Jion 引擎建立风险类型与跨表数据的映射关系, 对光伏项目投资风险进行实时监测。基于以上步骤, 本研究建立大数据应用监测模型, 该模型的监测过程是动态的, 监测结果是实时的。

3　大数据监测指标甄选方案设计

3.1　运用 CHAID 决策树归纳多维监测指标组合的方案设计

CHAID 决策树(CHi-square Automatic Interaction Detection Decision Tree)具有监督式的特征萃取与描述功能, 其基本思想是从训练数据集中自发地构造决策规则, 进而对其他数据集进行分类。决策树中的每个非叶子节点代表一个特征属性, 每个分枝表示这个特征属性的输出值, 每个叶子节点储存一个类别。规则的形成是从根节点开始, 对待分类项的特征属性进行测试, 并按照 CHAID 算法的卡方检验结果选择分支, 一直到达叶子节点, 最后将叶子节点储存的类别作为

决策结果。由于 CHAID 算法是以卡方检验结果作为分枝准则，因此不用考虑决策树修剪[18]。

以 Solarbao 平台光伏项目基本信息为基础，本研究通过 CHAID 决策树实时、动态地归纳多维监测指标组合。根据图 3，基本信息指 Solarbao 平台可投资项目列表界面中的项目信息。本研究定义其为基本指标，包含投资该项目的年化收益、产品单价、投资锁定期和返还利息方式。



图 3　Solarbao 平台光伏项目基本信息示例

每个基本指标包含 2 类属性，代表决策树分支；项目投资风险类型表示叶子节点，包含低风险型、中等风险型和高风险型，如表 2 所示：

表 2　基本指标及属性

基本指标	基本指标属性
(K1)年化收益	(K11)≤7% (K12)>7%
(K2)产品单价	(K21)≤2000 元 (K22)>2000 元
(K3)投资锁定期	(K31)≤90 天 (K32)>90 天
(K4)返还利息方式	(K41)到期返本返息 (K42)按月返息

表 3 是光伏项目投资风险专家判断 A 问卷的样例，问卷结果将被用于归纳多维监测指标组合。受访专家若认为一个年化收益不高于 7%、产品单价不高于 2 000 元、投资锁定期不超过 90 天、返还利息方式为到期返本返息的光伏项目为低风险型，则可在问卷中填写“LR”。

表 3　Solarbao 平台光伏项目投资风险专家判断 A 问卷

编号	年化收益	产品单价	投资锁定期	返还利息方式	项目投资风险类别判断
1	≤7%	≤2000 元	≤90 天	到期返本返息	LR
2	>7%	≤2000 元	>90 天	按月返息	MR
3	≤7%	>2000 元	≤90 天	到期返本返息	LR
4	>7%	>2000 元	≤90 天	到期返本返息	HR
5	≤7%	≤2000 元	>90 天	到期返本返息	LR
6	>7%	>2000 元	≤90 天	按月返息	MR
7	≤7%	>2000 元	>90 天	到期返本返息	MR
8	≤7%	>2000 元	>90 天	按月返息	LR
9	>7%	≤2000 元	≤90 天	到期返本返息	MR
10	>7%	>2000 元	>90 天	按月返息	HR
11	≤7%	≤2000 元	≤90 天	按月返息	LR
12	≤7%	≤2000 元	>90 天	按月返息	LR
13	>7%	≤2000 元	>90 天	到期返本返息	HR
14	≤7%	>2000 元	≤90 天	按月返息	MR
15	>7%	>2000 元	>90 天	到期返本返息	HR
16	>7%	≤2000 元	≤90 天	按月返息	LR

(注：LR 为低风险型；MR 为中等风险型；HR 为高风险型。“到期返本返息”是指在项目投资到期时返还投资者本金和全部利息；“按月返息”是指在项目投资存续期内每月返还投资者当月利息。)

3.2　运用 R-Q 型因子分析提炼关键监测指标的方案设计

R-Q 型因子分析(R-Q Mode Factor Analysis)是 R 型和 Q 型因子分析相结合的一种多元统计分析方法，由于 R 型因子分析和 Q 型因子分析具有对偶关系，因此变量点和样品点可投影于同一因子空间，且样品点

的类型能方便地被邻近的变量点来解释。两类因子分析之间存在的对偶关系在数学上已经被证明[19-20]，则通过正交变换即可从 R 型结果中推得 Q 型结果，其数学处理过程就是对应分析(Correspondence Analysis)。在同一标度的因子空间里，对应分析图形上样本点与变量点的邻近程度表示该变量因子对样本的解释程度。

以 Solarbao 平台光伏项目基本信息和专业信息为基础，本研究运用 R-Q 型因子分析实时、动态地提炼识别项目投资风险的关键指标。在图 4 示例中，专业信息是指点击 Solarbao 平台可投资项目列表中任一家项目后出现的融资信息，本研究定义其为专业指标，包括光伏项目运营商信息、项目建设进度表、相关资质、项目安全保障情况等；变量点由上述基本指标和专业指标共同组成，如表 4 所示。

项目简介　收益说明　相关资质　安全保障　相关问题

一．产品介绍

产品名称	金桔358号	发售价格	16.8元	固定收益	6.8%
发售数量	3000片	锁定期	90天	持有年限	90天
预期产生碳豆	0	付租方式	到期一次性还本付息		
销售期	2016.05.28	正常回购日	2016.08.26	合同到期日	2016.08.26
单个用户购买上限	50	付租日	2016.08.26	起租日	产品购买次日起租
融资租赁服务提供商	江苏绿能宝融资租赁有限公司		承租人	美桔新能源科技（苏州）有限公司	

二．承租人基本信息介绍

承租人介绍	美桔新能源科技（苏州）有限公司，成立于2015年01月08日，注册资本1000万美元，公司主要从事以电脑设计方式从事太阳能储能技术和节能技术的研发；节能领域技术服务、技术咨询；从事电子产品、太阳能储能产品、充电桩、光伏产品及原料、配套设备的批发、进出口、佣金代理（拍卖除外）及相关业务。	
承租人资质展示	请见“相关资质”	
所需租赁物信息及数量	多晶电池片A级	3000片
承租人租金支付方式	到期一次性还本付息	
融资租赁期限	2016.05.28——2016.08.26	
租赁物所有权的转移方式	承租人到期回购。	
担保方式	无	

图 4　Solarbao 平台光伏项目专业信息示例

表 4　基本指标、专业指标及属性

专业指标	专业指标属性
(A1) 投资年化收益较高	＞7%
(A2) 融资产品单价较高	＞2000 元
(A3) 投资锁定期较长	＞90 天
(A4) 返还利息频率较高	按月或季度返息
(A5) 项目总投资期较长	＞2 年
(A6) 总体融资规模较大	＞100 万
(A7) 承租企业成立年限较长	＞3 年
(A8) 承租企业公布的资质、合同齐全	公开了资质和合同复印件
(A9) 设备技术较为先进	光电转化率＞25%
(A10) 项目存续期预计年均发电量较大	＞1000 万度/年
(A11) 度电补贴占售电价比重较大	＞70%
(A12) 项目公布的实地照片较丰富	＞3 张

样本点即项目投资风险类型，包括低风险型、中等风险型和高风险型；表 5 是光伏项目投资风险专家判断 B 问卷的样例，问卷结果将被用于炼识别项目投资风险的关键指标。受访专家若认为 A1 指标“投资年化收益较高”符合中等风险型和高风险型光伏项目的特点，则可在对应的选项中打勾。

表 5　Solarbao 平台光伏项目投资风险专家判断 B 问卷

评价项目投资风险类型指标	低风险	中等风险	高风险
A1　投资年化收益较高		√	√
A2　融资产品单价较高		√	
A3　投资锁定期较长			√
A4　返还利息频率较高	√	√	
A5　项目总投资期较长		√	√
A6　总体融资规模较大	√		
A7　承租公司成立年限较长	√		
A8　承租公司公布的资质较齐全	√	√	
A9　设备技术较为先进			√
A10　项目存续期年均发电量较大	√	√	
A11　度电补贴占售电价比重较大	√		
A12　项目公布的实地照片丰富	√	√	

(注: “投资锁定期”是指投资者在确认投资后不得向第三方转让投资份额的期限; “承租公司公布的资质”包括公司营业执照、税务登记证、银行开户许可证、机构信用代码证、投资项目备案证等。)

在进行问卷调查时，若电力领域专家对问卷 A 和问卷 B 中专业金融术语存在内涵理解困难，问卷发放人员将向其进行详细解释，以确保电力领域专家对相关术语理解的准确度。基于问卷调查，本文获取专家对项目投资风险的判断结果和统计数据，并在此基础上采用 CHAID 决策树和 R-Q 型因子分析对数据进行分析，从专家判断中提炼监测光伏项目投资风险的多维监测指标组合和关键指标，为项目投资风险大数据监测指标甄选提供支持。

4　实证分析与结果讨论

共向受访专家①发放了 85 份 A 问卷和 32 份 B 问卷。回收有效问卷 A 68 份，有效问卷回收率为 80%，因每份 A 问卷包含 16 个判断结果，故 CHAID 决策树建模样本的数量为 1 088 个；回收有效问卷 B 30 份，有效问卷回收率约为 93.8%，因每份 B 问卷包含 12 个

①受访专家指在金融领域和电力领域工作 6 年以上或拥有相关领域博士学位的专业学者。

判断结果，故 R-Q 型因子分析的样本数量为 360 个。

4.1 基于 CHAID 决策树归纳多维监测指标组合

CHAID 决策树的总体分类正确率为 66.9%，其中，低风险型项目分类正确率为 84.2%，中等风险型项目分类正确率为 45%，高风险型项目分类正确率为 70.2%，该决策树共产生 8 条决策规则，如表 6 所示：

表 6　CHAID 决策树输出规则

编号	先行条件(if)				结果(then)
	年化收益	产品单价	投资锁定期	返还利息方式	
1	≤7%	≤2000 元或＞2000 元	≤90 天	按月返息	低风险型(89.0%)
2	＞7%	＞2000 元	＞90 天	到期返本返息　或按月返息	高风险型(77.9%)
3	≤7%	≤2000 元或＞2000 元	≤90 天	到期返本返息	低风险型(69.6%)
4	＞7%	≤2000 元	＞90 天	到期返本返息　或按月返息	中等风险型(69.4%)
5	≤7%	≤2000 元或＞2000 元	＞90 天	按月返息	低风险型(65.4%)
6	＞7%	＞2000 元	≤90 天	到期返本返息　或按月返息	中等风险型(63.2%)
7	＞7%	≤2000 元	≤90 天	到期返本返息　或按月返息	低风险型(62.0%)
8	≤7%	≤2000 元或＞2000 元	＞90 天	到期返本返息	高风险型(39.7%)

在表 6 中，8 条决策规则的表述方式是“if-then”形式。例如，第 1 条决策规则表示：若光伏发电项目的年化收益不高于 7%、产品单价不高于 2 000 元(或高于 2 000 元)、投资锁定期不超过 90 天、返还利息方式为按月返息，该项目被认为是低风险型；第 2 条决策规则表示：若光伏发电项目的年化收益高于 7%、产品单价高于 2 000 元、投资锁定期超过 90 天、返还利息方式为到期返本返息(或按月返息)，则该项目被认为是高风险型。在以上决策规则中，各投资风险类型可由 4 个监测指标表征，故本研究基于 CHAID 决策树得到 8 条多维监测指标组合。

4.2 基于 R-Q 型因子分析提炼关键监测指标

表 7 是 B 问卷的专家判断汇总结果和 R-Q 型因子分析结果。专家判断汇总结果是指各指标对应项目投资风险类型被专家勾选的数量占回收问卷总数的百分比；R-Q 型因子分析结果是指监测指标与项目投资风险类型的相关度统计，用卡方值衡量。

表 7　专家判断汇总结果与 R-Q 型因子分析结果

监测指标	专家判断汇总结果			R-Q 型因子分析结果(卡方值)		
	低风险型	中等风险型	高风险型	低风险型	中等风险型	高风险型
A1	21.9%	46.9%	87.5%	–3.3	–0.3	3.7
A2	12.5%	40.6%	75.0%	–3.6	0.0	3.7
A3	18.8%	56.3%	78.1%	–3.6	0.8	2.9
A4	81.3%	46.9%	21.9%	2.8	–0.1	–2.8
A5	21.9%	40.6%	65.6%	–2.6	0.0	2.6
A6	40.6%	50.0%	59.4%	–1.3	0.2	1.1
A7	75.0%	40.6%	28.1%	2.5	–0.6	–2.0
A8	71.9%	43.8%	25.0%	2.3	–0.1	–2.2
A9	37.5%	53.1%	65.6%	–1.8	0.3	1.5
A10	65.6%	46.9%	15.6%	2.1	0.9	–3.0
A11	84.4%	34.4%	12.5%	4.1	–0.8	–3.3
A12	75.0%	43.8%	25.0%	2.5	–0.2	–2.3

基于 R-Q 型因子分析结果，项目投资风险类型与监测指标对应分析图如图 5 所示：

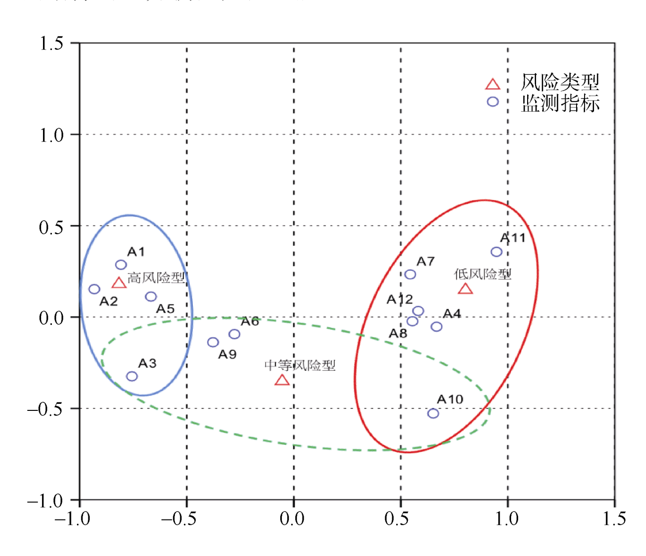


图 5　项目投资风险类型与监测指标对应分析

(1) 低风险型项目与 A4、A7、A8、A10、A11 和 A12 指标组成第一类聚类，说明若项目返还利息频率较高、承租公司成立年限较长、承租公司公布的资质较齐全、项目存续期年均发电量较大、度电补贴占售电价比重较大、公布的实地照片丰富，则该项目属于低风险型；

(2) 高风险项目与 A1、A2、A3 和 A5 指标组成第二类聚类，说明若项目投资年化收益较高、融资产品单价较高、投资锁定期较长、项目总投资期较长，则该项目属于高风险型。

(3) 在中等风险型项目风险判断方面，根据卡方统计显著性，A3 和 A10 指标对应于中等风险型的卡方值分别为 0.8 和 0.9，A6 和 A9 指标对应于中等风险型的卡方值也为正，故本研究将这 4 项指标归为中等风险型项目的辅助监测指标。

总结以上结果还可以发现：

(1) 若项目投资年化收益较高、融资产品单价较高、投资锁定期较长，则该项目被认为具有高风险，这与 CHAID 决策树归纳的第 2 条决策规则相符；

(2) 若项目返还利息频率较高，则该项目被认为属于低风险型，这与 CHAID 决策树归纳的第 1 和第 5 条规则相符。这说明 CHAID 决策树模型实证结果中有 3 条规则在 R-Q 型因子分析实证结果中得到了验证，表明该甄选方案具有可靠性。

5　结　语

在构建光伏项目投资风险监测模型的过程中，甄选面向互联网金融平台的大数据应用监测指标是其核心难点。本研究构建了实时动态监测光伏项目投资风险的大数据监测模型，为满足模型对指标采集的要求，进而提出大数据监测指标甄选方案：在整合 Solarbao 平台多源异构数据基础上，以专家判断为项目投资风险分析依据，运用 CHAID 决策树归纳多维监测指标组合，并运用 R-Q 型因子分析方法提炼识别投资风险的关键指标。研究得到 8 条“if-then”形式的光伏项目投资风险监测指标组合和 10 项识别投资风险的关键指标，表明该甄选方案具有可行性，符合监测模型对指标采集的要求。在 CHAID 决策树模型实证结果中有 3 条规则与 R-Q 型因子分析实证结果实现了相互验证，表明该甄选方案具有可靠性。研究的局限在于 R-Q 型因子分析中的专业指标有待进一步细分并形成动态更新机制。

目前，电力新能源领域互联网金融在光伏项目融资过程中扮演着越来越重要的角色，但其背后积累的融资主体信用风险也不容忽视，加强光伏项目投资风险监测成为推进该领域互联网金融持续健康发展的客观需要和当务之急。本文所提出的光伏项目投资风险大数据监测指标甄选方案符合风险监测的实际应用需要，为投资者评估光伏项目风险、平台筛选合适项目以及监管部门排查该领域系统性风险提供了一种手段，具有理论意义和实践价值。

参考文献：

[1] Scholtens B. Finance as a Driver of Corporate Social Responsibility [J]. Journal of Business Ethics, 2006, 68(1): 19-33.

[2] Climent F, Soriano P. Green and Good? The Investment Performance of US Environmental Mutual Funds [J]. Journal of Business Ethics, 2011, 103(2): 275-287.

[3] Graham A, Maher J J, Northcut W D. Environmental Liability Information and Bond Ratings [J]. Journal of Accounting Auditing & Finance, 2001, 16(2): 93-116.

[4] Thomas S, Repetto R, Dias D. Integrated Environmental and Financial Performance Metrics for Investment Analysis and Portfolio Management [J]. Corporate Governance: An International Review, 2007, 15(3): 421-426.

[5] Pope D G, Sydnor J R, What's in a Picture? Evidence of Discrimination from Prosper.com [J]. Journal of Human Resources, 2008, 46(1): 53-92.
[6] Duarte J, Siegel S, Young L. Trust and Credit: The Role of Appearance in Peer-to-Peer Lending [J]. Review of Financial Studies, 2012, 25(8): 2455-2484.
[7] 栾春玉，代榕家. 网络金融信息生态治理模式与管控对策研究[J]. 情报科学，2015, 33(5): 48-52. (Luan Chunyu, Dai Rongjia. Study on the Ecological Management Mode and Control Strategy of Internet Financial Information [J]. Information Science, 2015, 33(5): 48-52.)
[8] 张立超，房俊民，唐钦能. 产业竞争情报预警工作中的风险识别研究：以我国光伏发电产业为例[J]. 情报理论与实践，2011, 34(10): 52-55. (Zhang Lichao, Fang Junming, Tang Qinneng. Research on Risk Identification in the Early Warning of Industry Competitive Intelligence: A Case Study of Photovoltaic Power Generation Industry in China [J]. Information Studies: Theory & Application, 2011, 34(10): 52-55.)
[9] 白洋. 面向大数据的电力设备状态监测信息聚合研究[D]. 昆明：昆明理工大学，2014. (Bai Yang. Research on Data Aggregation of Power Equipment Condition Monitoring Based on Big Data [D]. Kunming: Kunming University of Science and Technology, 2014.)
[10] 郑海雁，金农，季聪，等. 电力用户用电数据分析技术及典型场景应用[J]. 电网技术，2015, 39(11): 3147-3152. (Zheng Haiyan, Jin Nong, Ji Cong, et al. Data Analysis Technology and Typical Application of Electric Power User [J]. Power System Technology, 2015, 39(11): 3147-3152.)
[11] Takabi H, Joshi J B D, Ahn G J. Security and Privacy Challenges in Cloud Computing Environments [J]. Security & Privacy IEEE, 2010, 8(6): 24-31.
[12] 路永和，曹利朝. 基于关联规则综合评价的图书推荐模型[J]. 现代图书情报技术，2011(2): 81-86. (Lu Yonghe, Cao Lizhao. Book Recommendation Model Based on Comprehensive Evaluation of Association Rules [J]. New Technology of Library and Information Service, 2011(2): 81-86.)
[13] Wang C H, Hsuesh O Z. A Novel Approach to Incorporate Customer Preference and Perception into Product Configuration: A Case Study on Smart Pads [J]. Computer Standards & Interfaces, 2013(35): 549-556.
[14] 孙霄凌，赵宇翔，朱庆华. 在线商品评论系统功能需求的Kano 模型分析——以我国主要购物网站为例[J]. 现代图书情报技术，2013(6): 76-84. (Sun Xiaoling, Zhao Yuxiang, Zhu Qinghua. Analyzing the Demand of Online Product Review System's Features Using Kano Model: An Empirical Study of Chinese Online Shops [J]. New Technology of Library and Information Service, 2013(6): 76-84.)
[15] Wang C H. Incorporating Customer Satisfaction into the Decision-Making Process of Product Configuration: A Fuzzy Kano Perspective [J]. International Journal of Production Research, 2013(22): 6651-6662.
[16] Nagamachi M. Kansei Engineering: A New Ergonomic Consumer-Oriented Technology for Product Development [J]. International Journal of Industrial Ergonomics, 1995, 15(1): 3-11.
[17] 程铁信，郭涛，祁昕. 决策树分类模型在工程项目评标风险预警中的应用[J]. 数理统计与管理，2010, 29(1): 122-128. (Cheng Tiexin, Guo Tao, Qi Xin. Application of Decision Tree Classification Model in Risk Early Warning of Engineering Project Evaluation [J]. Journal of Applied Statistics and Management, 2010, 29(1): 122-128.)
[18] Dey P K. Project Risk Management Using Multiple Criteria Decision Making Technique and Decision Tree Analysis: A Case Study of Indian Oil Refinery [J]. Production Planning & Control, 2012, 35(3): 1-19.
[19] Walden J, Smith J P, Dackombe R V. The Use of Simultaneous R-Q Mode Factor Analysis as a Tool for Assisting Interpretation of Mineral Magnetic Data [J]. Mathematical Geology, 1992, 24(3): 227-247.
[20] Murtagh F. The Correspondence Analysis Platform for Uncovering Deep Structure in Data and Information [J]. The Computer Journal, 2010, 53(3): 304-315.

作者贡献声明:

杨旸：细化研究命题和研究思路，起草论文，清洗、加工和分析数据；
林辉：提出研究命题，设计研究方案，数据采集，进行实验，修改论文；
胡广伟：优化研究方案，设计实验流程，论文修改及最终版本修订。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: hugw@nju.edu.cn。
[1] 杨旸，林辉，胡广伟. IndexSelection.rar. 光伏项目风险监测指标甄选实验数据.

[2] 杨旸, 林辉, 胡广伟. SolarbaoDataExtraction.rar. Solarbao 平台数据抓取程序包.
[3] 杨旸, 林辉, 胡广伟. QuestionnaireA.rar. A 问卷调研统计结果.
[4] 杨旸, 林辉, 胡广伟. QuestionnaireB.rar. B 问卷调研统计结果.
[5] 杨旸, 林辉, 胡广伟. DecisionTree.spv. CHAID 决策树模型实验输出.
[6] 杨旸, 林辉, 胡广伟. CorrespondenceAnalysis.spv. R-Q 型因子分析模型实验输出.

收稿日期: 2016-07-25
收修改稿日期: 2016-08-29

Detecting Investment Risks of Photovoltaic Projects with Big Data: Case Study of Solarbao.com

Yang Yang[1] Lin Hui[1] Hu Guangwei[2]
[1](School of Business, Nanjing University, Nanjing 210093, China)
[2](School of Information Management, Nanjing University, Nanjing 210093, China)

Abstract: [**Objective**] This research proposes a selection scheme for the big data application to monitor the Internet financial platforms, which is verified by the real world cases. [**Methods**] First, we adopted a big data model to integrate multi-source heterogeneous data from the Solarbao platform. Second, we utilized the CHAID decision tree to summarize multi-dimensional monitoring indicators based on analysis of each project's investment risks. Finally, we employed the R-Q factor analysis method to extract the key investment risks. [**Results**] We got 8 indicators to track the investment risks, which could be identified by the other 10 indicators for the photovoltaic projects. [**Limitations**] More research needs to be done with indicators of the R-Q factor analysis, which also requires a dynamic update mechanism. [**Conclusions**] The proposed scheme could help investors assess the risks of individual projects and then select the appropriate ones. It will also support the risk management work of the regulatory agencies.

Keywords: Big data monitoring index Photovoltaic project Investment risk CHAID decision tree R-Q mode factor analysis